Daniel E. Geer, Jr.
In-Q-Tel

Dale Peterson
S4 Events

Failure as Design

Cybersecurity is perhaps best understood as a constraint-based design problem: what are the failure modes the design team is unwilling to accept? The countervailing argument—that cybersecurity design is all about enablement—is less utile. Security failures that are unacceptable seem often to trace to a constraint ignored in design stages. This being a column, let's jump straight to current issues.

Heretofore, well-designed networked applications have always included detection of, and compensation for, outright network failure. ("If disconnected, then revert to Operating Plan B.") Attending to the constraint of unreliable components is why the whole of TCP/IP is what it is (tolerant of random faults).

Looking forward, the design envelope in the 5G world seems not to include the constraint of detection and compensation for network failure, but rather to assume never-off connectivity. If the constraint of unreliable communications is not within the design space of 5G devices, there are profound downstream implications not limited to mere denial of service.

Of course, there is a difference between inadvertently failing to account for a constraint and only appearing to do so inadvertently. The pinnacle constraint in cybersecurity is the constraint of sentient opponents, and, as such, it is (always) possible that a flaw appearing to trace to an unacknowledged technical constraint is/was no accident. In their June 2019 paper,[1] Finite State made a number of claims of exactly this sort:

- Huawei engineers disguised known unsafe functions (such as memcpy) as the "safe" version (memcpy_s) by creating wrapper functions with the "safe" name but none of the safety checks. This leads to thousands of vulnerable conditions in their code.
- There are several million calls into unsafe functions. Huawei engineers choose the "safe" option of these functions less than 17% of the time, despite the fact that these functions improve security and have existed for over a decade.

So with respect to the constraint to not use unsafe functions, was violation of that constraint simply that it was neglectfully unrecognized, or does this trace to sentient opponents? Continuing…

- 76 instances of firmware where the device was, by default, configured such that a root user with a hard-coded password could log in over the SSH protocol, providing for default backdoor access.
- eight different firmware images were found to have precomputed authorized_keys hard-coded into the firmware, enabling backdoor access to the holder of the private key.
- 424 different firmware images contained hardcoded private SSH keys, which can enable a man-in-the-middle to manipulate and/or decrypt traffic going to the device.

So is the constraint to only allow remote access via securable mechanisms a constraint that was neglectfully unrecognized, or does its violation trace to sentient opponents?

One can argue, and some do, that flaws of the sort Finite State found are typical of large corporations engaged in races to markets, markets whose structure inherently confers near-insurmountable first-mover advantages. That would imply that such findings as those of Finite State do not trace to sentient opponentry but rather to something more banal.

Even if it is the banal alternative, for policy people is that a distinction but not a difference? In medicine, one would say that health care stops and research begins when further refinement of an obscure diagnosis does not lead to a different treatment or, in the lingo of medicine, where there is "no therapeutic difference." Put differently, what constraints do policy makers operate under that technical folks do not? How do they rightly factor in the ambiguity of sentient opponents who are sentient enough to appear to inadvertently distribute vulnerabilities?

Combining the 5G always-on assumption (the absence of a design constraint for

Digital Object Identifier 10.1109/MSEC.2019.2936706
Date of current version: 29 October 2019

continued on p. 89

 1540-7993/19©2019IEEE

 continued from p. 90

unreliable communications) with the findings of Finite State (that 5G communications as fielded by the world market leader can be made to fail through known flaws), what do we have? An embedded flaw base in a must-be-on, society-wide dependency, salted with the stockpiling of tools against such up and down the skill ladder, is hardly an issue to be dismissed. Whether it can be usefully anticipated in ways far better technically than our current ways is at question. Preservation of an operational base that does not share 5G common-mode failures certainly leads to hard choices at the policy level.

Again, the premise of always-ready connectivity is profound, but can its contribution to risk be overstated? In discussions around 5G and industrial control systems (ICS), at present there may be a general overestimation of the impact of an outage of ICS—that is to say there are real settings where an outage would not be a high-consequence event.

For example, many manufacturers' operations can still be run manually, and for various reasons they do so many times a year (when something breaks). This ability is decreasing, but it is still common—even in power plants, a lot of the balance of plant systems, e.g., coal handling, water systems, etc., can be run manually; only the turbine control system cannot. Similarly, most manufacturers are not running three shifts, seven days a week, so if it took them three days to recover from a cybercaused outage, they would just convert to three shifts and be caught up in a week.

Often the cost of an outage just isn't that great. For an electric utility, being unable to produce power in a plant is counterbalanced by their ability to buy power from elsewhere. Even a month-long unit outage doesn't come close to the kind of loss that would be addressed at the executive level (although executive management is highly concerned about reputational impacts). High financial loss is when there is equipment or site damage that requires engineering and automation skills in addition to access. Yes, electric transmission is a case where an outage for even hours is a high-consequence event, but such cases are a small minority of the ICS.

That being said, Supervisory Control and Data Acquisition (SCADA) systems are increasingly dependent on "the Internet" as provided by the carriers. While SCADA components may well be segmented away from the general Internet, if the carrier is unable to provide Internet services, then it is likely SCADA communication would also go down. Note that the combination of the much better prices and the carriers no longer offering dial-up or leased-line services is moving pipelines, water distribution, and other SCADA to mobile data services. This is even working its way into the electric grid beginning at the pole-top systems.

As a forecasting question, when do the systems that monitor and control a process in a building or site come to rely on the cloud and wide area communication? Will we start to see this as soon as the next one to three years? Doing some or all of operations and maintenance remotely is heresy today, but what an operator actually does is ideal for machine learning (and there are long historical records to train on). Perhaps remote control won't happen for a long time for the very most essential systems, but otherwise the ICS product vendors, integrators, and others will soon do a high percentage of the work that is today done on site from off-site tomorrow. At that point, the putative constraint of occasional communications failure becomes very important.

It has been shown that a significant vendor has a very clever and hard-to-find backdoor in its systems. It could be there for support reasons, but it is also a great example of what a nation state would like to have, something in their back pocket in case it was ever needed. Does the motivation actually matter and, if so, to whom?

Summing up, is "cybernationalism" an inevitable, organic consequence of a society realizing that it is unable to withstand outages? Are we moving to a world where you only buy systems from "your" team? There are many U.S.-made critical infrastructure components used in China; do these go away? Does it become the flip side of high-end weapon systems sales where you only buy from entities close enough to sell you their good stuff? What is the constraint space for policy makers now (or then)? As Norwegian Olav Lysne points out, for a smaller country the only policy choice is whose vulnerabilities, both inadvertent and intentional, do you adopt.

The root cause of risk is dependence, and particularly a dependence on the stability of system state. If you define a state of security as the absence of unmitigatable surprise, how do you tease out the constraints that ensure your surprises are ones you can mitigate? Is 5 G incompatible with critical dependencies due to its design constraints, or is that mitigatable by cybernationalism? Is the ICS where such questions soon come to a head? ■

Reference

1. "Finite State Supply Chain Assessment: Huawei Technologies Co., Ltd," finitestate.io/wp-content/uploads/2019/06/Finite-State-SCA1-Final.pdf

Daniel E. Geer, Jr. is the chief information security officer of In-QTel. Contact him at dan@geer.org.

Dale Peterson is the founder and project chair of S4, an advanced ICS security conference held every January in Miami South Beach. He can be reached through https://dale-peterson.com/contact/.